

AraQTL manual

AraQTL- Workbench and Archive for Natural Variation Research in *Arabidopsis thaliana*.

Harm Nijveen^{1,2*}, Wilco Ligterink², Joost J.B. Keurentjes³, Olivier Loudet⁴, Jiao Long^{1,5}, Henk W. Hilhorst², Dick de Ridder¹, Jan E. Kammenga⁵ & Basten L. Snoek^{5*}

¹Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands

²Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands

³Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

⁴Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, 78000 Versailles, France

⁵Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands

*Corresponding authors

Email: Harm.nijveen@wur.nl; Basten.snoek@wur.nl

Genetical genomics studies uncover genome-wide genetic interactions between genes and their transcriptional regulators. High-throughput measuring of gene expression in recombinant inbred line populations enabled the investigation of the genetic architecture of gene expression variation. This has the potential to enrich the understanding of the molecular mechanisms affected by and underlying natural variation. Moreover, it is contributing to the systems biology of natural variation, as a substantial number of experiments have resulted in a valuable amount of interconnectable phenotypic, molecular and genotypic data. For *Arabidopsis thaliana* a number of genetical genomics studies have been published, uncovering many expression quantitative trait loci (eQTLs). Yet, their complex results are not easily accessible to the plant research community, leaving most of the valuable genetic interactions unexplored as cross-analysis of these studies is a major effort. We address this with AraQTL (<http://www.bioinformatics.nl/AraQTL/>), an easily accessible workbench and database for comparative analysis and meta-analysis of all published *Arabidopsis* eQTL datasets. AraQTL provides a workbench for comparing, re-using and extending upon the results of these experiments. For example, one can easily screen a physical region for specific local-eQTLs that could harbour candidate genes for phenotypic QTL or detect gene-by- environment interactions by comparing eQTLs under different conditions.

Index

1. Getting started using the homepage and navigation bar
2. Genome-wide overview by “Experiment” function
3. Comparing gene specific eQTL profiles from multiple experiments
4. The “Locus” function shows co-locating eQTLs
5. The “Correlation” function shows correlated eQTL profiles

1) Getting started using the homepage and navigation bar

The homepage of AraQTL offers an overview of the studies and dataset currently hosted and the navigation tools for investigation and comparison (**Figure 1**). The search box can be used to find the eQTL profiles of one or more genes in one or all experiments. Moreover GO terms can be used to find the eQTL profiles of all genes annotated with that GO term. Any query term not directly linked to a gene ID or GO term will report the genes with that query-term in their description.

Experiments to be visualised can be pre-selected from the experiment overview table. This table also offers some basic information such as population used, plant parts and original publication. The publication titles are linked to their PubMed pages for easy access and the data hosted in AraQTL can be downloaded in flat text format or accessed through programming languages like R.

The screenshot shows the AraQTL homepage with the following features highlighted by numbered callouts:

- ⑦ AraQTL Home (Navigation bar)
- ⑧ Experiment (Navigation bar)
- ⑨ Correlation (Navigation bar)
- ⑩ Locus (Navigation bar)
- ⑪ Help (Navigation bar)
- ⑫ Examples (Navigation bar)
- ① Search box with placeholder text: "type your query, i.e. AT1G01150, or GO:0004674, or ..."
- ⑬ Help links: "What is AraQTL?", "What is an eQTL?", "How can I find information about my gene?"
- ② Choose an experiment: (Table header)
- ③ Cross (Table header)
- ④ Plant parts (Table header)
- ⑤ Publication (Table header)
- ⑥ Download (Table header)

② Choose an experiment:	③ Cross	④ Plant parts	⑤ Publication	⑥ Download
* Joosen_etal_2012	Bay x Sha	Seeds	Visualizing the genetic landscape of Arabidopsis seed per...	lod.txt
	Rav v		Identifying genotype-	

Figure 1: AraQTL homepage. From this page different investigations can be started. It contains several key features: ① The main search box, from which genes, markers, gene descriptions and GO terms can be found. Searching for a single gene shows the eQTL profile for that gene for the selected experiment or all experiments when selected. Below this, the summary table lists datasets currently in AraQTL which consists of: ② Data set selection, ③ Recombinant inbred population used, ④ Plant parts from which gene expression was measured, ⑤ Title and link to the publication the dataset was originally generated for and ⑥ a link to download the dataset with the eQTL profiles in simple txt format. On the top of the page a navigation bar can be found from which several selections can be made: ⑦ the link to the homepage, ⑧ a link to a genome wide overview by cis/trans plot per experiment, ⑨ a link to for eQTL profile correlation investigation, ⑩ a link to find all eQTLs at a specific locus, ⑪ a link to help features like, About, the manual and FAQ and ⑫ a link to a page with clickable graphical examples of the different investigations. At ⑬ quick links to important information can be found.

The navigation bar provides the main form of interactivity (**Figure 2**) and can be used for selection of graphical overviews, investigations and information. From left to right: the “AraQTL Home” button enables quick return to the home page and search box. The “Experiment” option can be used to go to a genome wide overview for each experiment. The “Correlation” option can be used to find genes with correlating eQTL profiles per experiment. The “Locus” option can be used to find all genes with an eQTL at a specific marker or genomic coordinate. “Help” offers information about AraQTL, a link to this manual and Frequently Asked Questions. Through “Examples” a visual and clickable overview of the main investigation options can be selected.

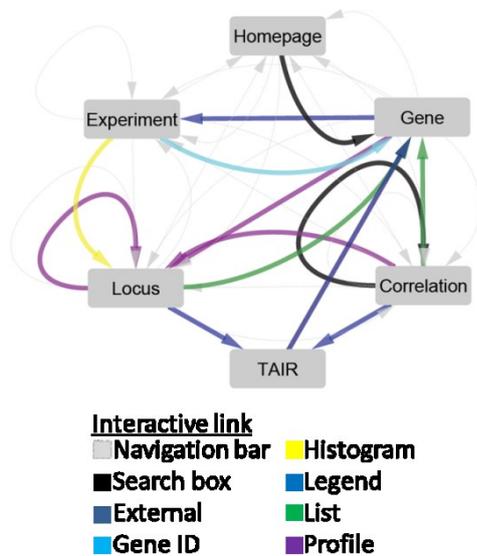


Figure 2: Contactmap of AraQTL. Legend shows the interactive links and which parts can be selected to go to further with an investigation. Dotted grey arrows show the links via the top navigation bar. Black arrows show connections via the search box. Dark blue lines represent the connections through the figure legend. Light blue arrows represent the links via the plot title. Yellow arrows show the connections via the eQTL distribution histogram. Purple arrows indicate the connections through the eQTL profile and green arrows show the connections by the gene list below the eQTL profile figures.

2) Genome-wide overview by “Experiment” function

To get a genome-wide overview of the distribution of eQTLs, a cis/trans Plot (**Figure 3**) is most frequently used. The position of each eQTL peak is plotted on the x-axis against the positions of the genes for which this eQTL peak was found. A default eQTL peak significance threshold per experiment at FDR of 0.05 is provided for quick visualisation and an option to manually adjust this threshold for further exploration. This Cis/Trans plot has been made interactive, allowing to select different experiments by a drop-down menu. Moreover the individual significant peaks (points in the Cis/Trans Plot) can be clicked to show the complete eQTL profile of the selected gene/peak. Finally, the total peak count per marker is indicated in the bar plot below the Cis/Trans plot. Individual bars/markers can be clicked to obtain a figure and list with genes sharing a peak at the selected marker.

In this plot, several patterns can be detected. The diagonal line shows the likely cis-eQTLs, which are those eQTLs mapping close to the position of the gene for which they have been found. Also vertical bands (“trans-bands”) can be observed, these contain many eQTLs each, which map to the same (small) genomic region in which they themselves are not physically located. These trans-band loci regulate the expression of many genes and likely contain one or more key regulators, which act on transcription either directly or indirectly (through some other phenotypes).

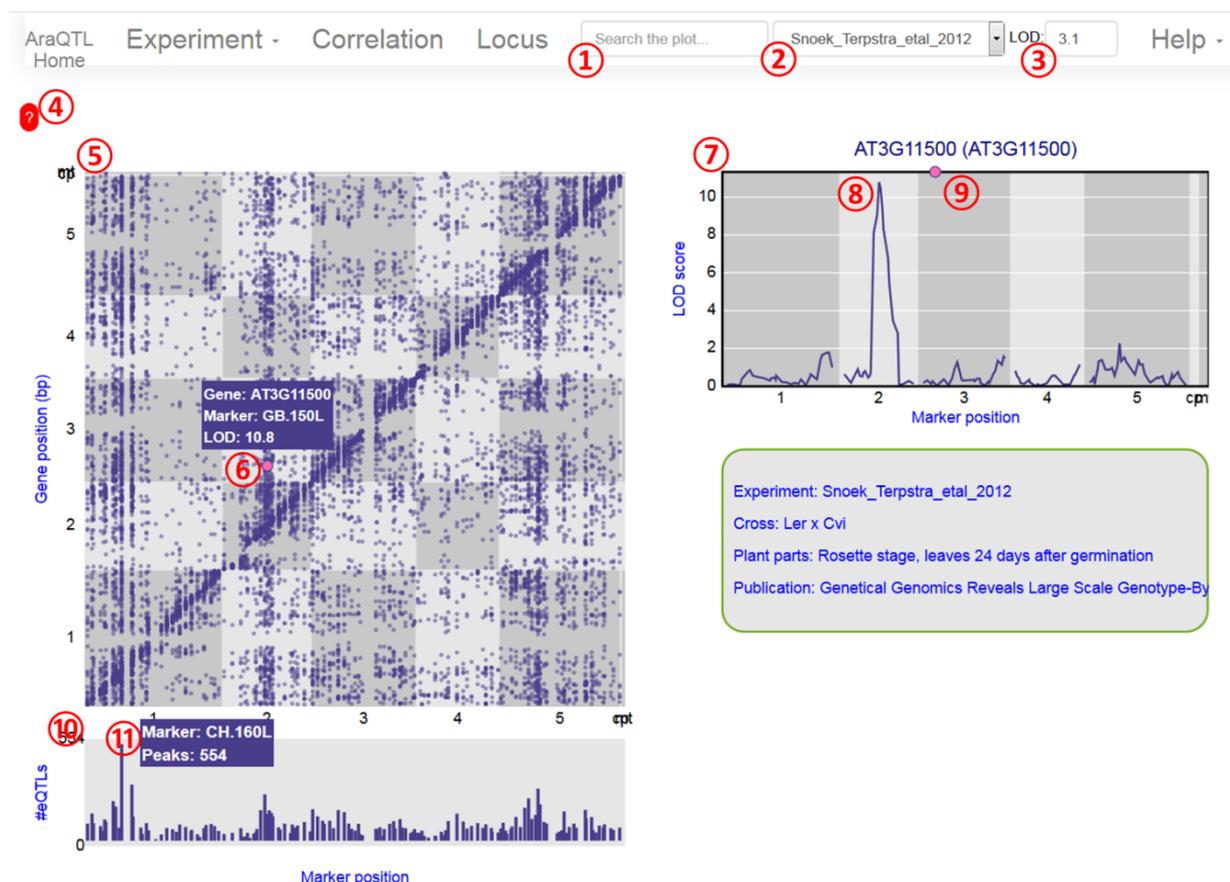


Figure 3: Cis/Trans plot. This page gives a genome wide overview per experiment of all the detected eQTLs. Individual genes can be found by using the search box (1) the experiment can be selected by the dropdown menu (2) and the significance threshold can be set at the navigation bar at the top of the page (3). Information about the genome-wide overview of eQTLs by the cis/trans plot can be found clicking the red indicator (4). In the cis/trans

plot ⑤ the genomic position of the eQTLs are plotted on the x-axis and the genomic position of the genes for which they have been found on the y-axis. The eQTLs spots ⑥ are interactive, indicating for which gene and at which marker the eQTL was found and what the LOD score of the peak was when the mouse hovers over a spot. When selected the full eQTL profile of that gene is shown in ⑦. In this graph the position of the gene ⑨ is shown by a pink dot at top of the panel. The graph is interactive ⑧ and shows the marker when the mouse hovers over the eQTL profile. Below this profile figure some information about the experiment is shown. In ⑩ a histogram is shown, visualising the distribution of eQTLs over the genome. The bars ⑪ of this histogram can be inspected to show the marker and exact number of eQTLs at that marker. When a bar is clicked a new page is opened with all the eQTL profiles of genes having a significant eQTL at that marker.

3) Comparing gene specific eQTL profiles from multiple experiments

In the “Gene eQTL profiles” window gene specific eQTL patterns between experiments can be compared (Figure 4). In this way experiment specific and co-occurring eQTLs can be easily recognised. Genes can be searched and selected by a search box. The figure itself is interactive and contains features facilitating further investigation. Points in the eQTL profiles can be clicked to open a new window showing genes with an eQTL at the clicked marker in the selected experiment. Experiment names in the legend can be moused-over to highlight a specific profile or clicked to go to the cis/trans plot for a genome-wide overview of the selected experiment. Some information about the gene can be found below the graph. From this type of investigation, comparing eQTL profiles from different experiments, can help deduce if the expression levels of a gene could be affected by interactions between genetic polymorphisms and environmental conditions. Moreover it shows those experiments for which gene expression variation exists that can be pinpointed to a specific locus. These experiments can then be selected for a more detailed look at the transcriptional regulation of the gene for which the eQTLs are visualised.

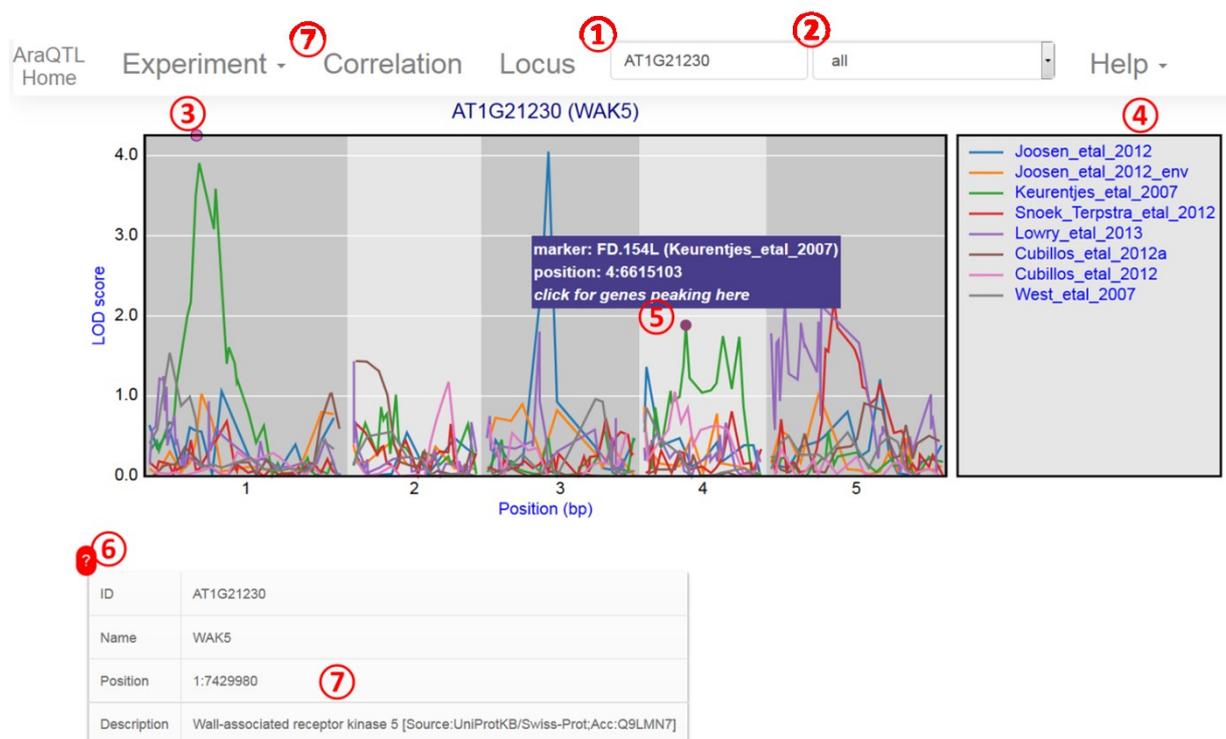


Figure 4: Multiple experiment plot. This page gives the eQTL profiles for the selected genes in all experiments in AraQTL. In this way the differences and similarities in eQTLs between the experiments can be studied. Overlapping eQTLs indicate regulatory regions active in multiple environments or populations, whereas unique eQTLs indicate regulators only active in a specific environment or population. At ① the gene of interest can be selected and at ② the experiment. At ③ the genomic position of the gene is shown by the pink dot. The legend on the right ④ can be moused-over to highlight specific eQTL profiles. The experiments can be clicked to go to the cis/trans plot for a genome-wide overview. The eQTL profiles can be moused-over ⑤ to get information about the marker, experiment and significance. When clicked all eQTL profiles with an eQTL at the clicked position and experiment are visualised. Information about the figure can be found by clicking ⑥. At ⑦ a table with information about the gene under investigation can be found below the eQTL profiles.

4) The “Locus” function shows co-locating eQTLs

To find genes with an eQTL at the same marker the “Locus” window can be used. The profile of the ten genes with the highest LOD score is shown in the figure and all genes with a peak above the threshold are listed below the figure (**Figure 5**). Experiment, marker, exclusion window and threshold can be selected to easily switch between experiments and positions. The exclusion window enables differentiation between potential-cis and trans eQTLs for further refinement of genes mapping to a certain locus. The figure is interactive, with the gene name in the legend linked to TAIR (www.arabidopsis.org) for up-to-date information. By clicking on the gene name in the table the eQTL profiles from all experiments for that gene will be visualised. From this type of investigation we can find which groups of genes share a common regulatory locus and under which conditions. This group of genes can now be studied in more detail, for example for a shared transcription factor binding site or receptor.

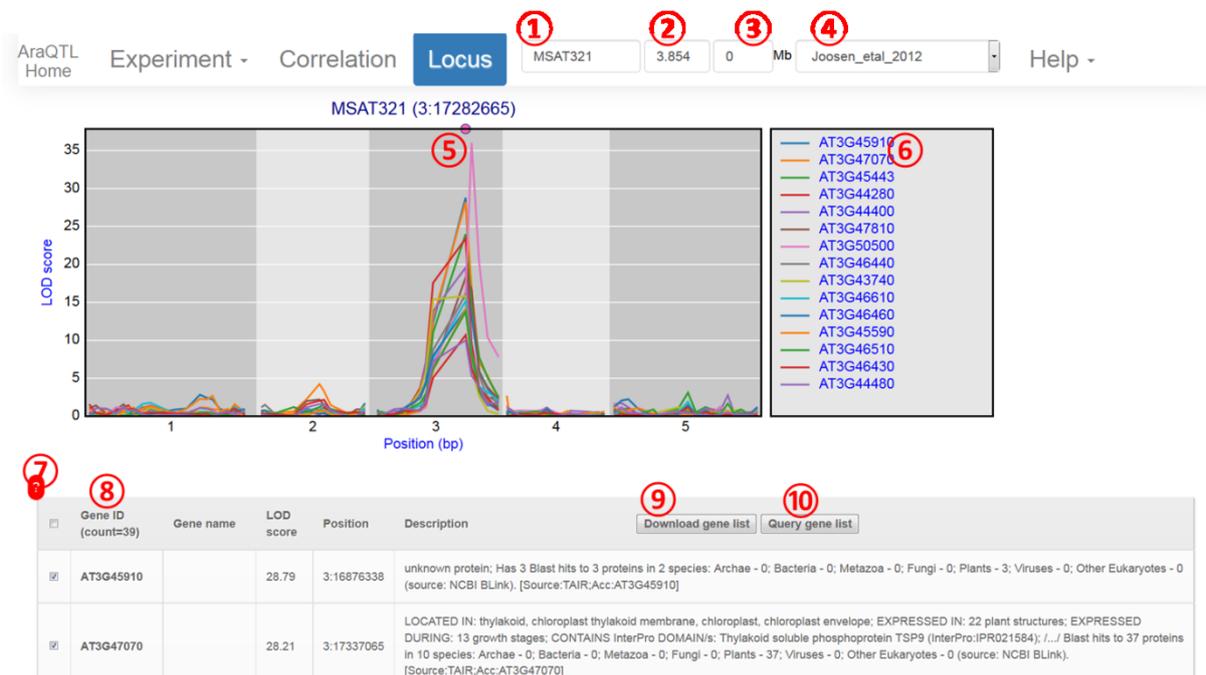


Figure 5: Co-locating eQTL. This page shows the eQTL profiles of genes with a co-locating eQTL at a specific locus. Several selections can be made from the navigation bar: ① the marker, ② the significance threshold, ③ gene position exclusion window and ④ the experiment. The selected marker is indicated as a pink dot ⑤ at the upper side of the plot. Furthermore the eQTL profiles can be moused-over for more information and clicked for further investigation or refinement. The legend ⑥ can be moused-over to highlight the eQTL profiles, clicking a gene ID will link to TAIR for more information about the gene. A list with all the genes with an eQTL at the selected locus can be found below the plot ⑦ from which the visualised eQTL profiles can be selected. The gene IDs at ⑧ can be clicked to show the eQTL profiles from all experiments. The list of genes can be downloaded using ⑨ for further investigations or sent to the query box ⑩ so the eQTL profiles of the selected genes can be studied in other experiments.

5) The “Correlation” function shows correlated eQTL profiles

Extending from co-location of eQTL peaks the “Correlating eQTL profiles” window can be used to find genes with a correlated eQTL profile. This shows genes which expression levels are controlled by (at least partly) co-localised eQTLs and hence have a comparable genetic architecture (**Figure 6**). This is more specific than what can be found with the co-locating eQTL function were genes are selected which share one regulatory locus. Interactivity is like described for previous pages so experiment, gene and correlation threshold can be selected from the navigation bar and specific peaks in the figures can be clicked to find genes sharing eQTLs at the selected marker.

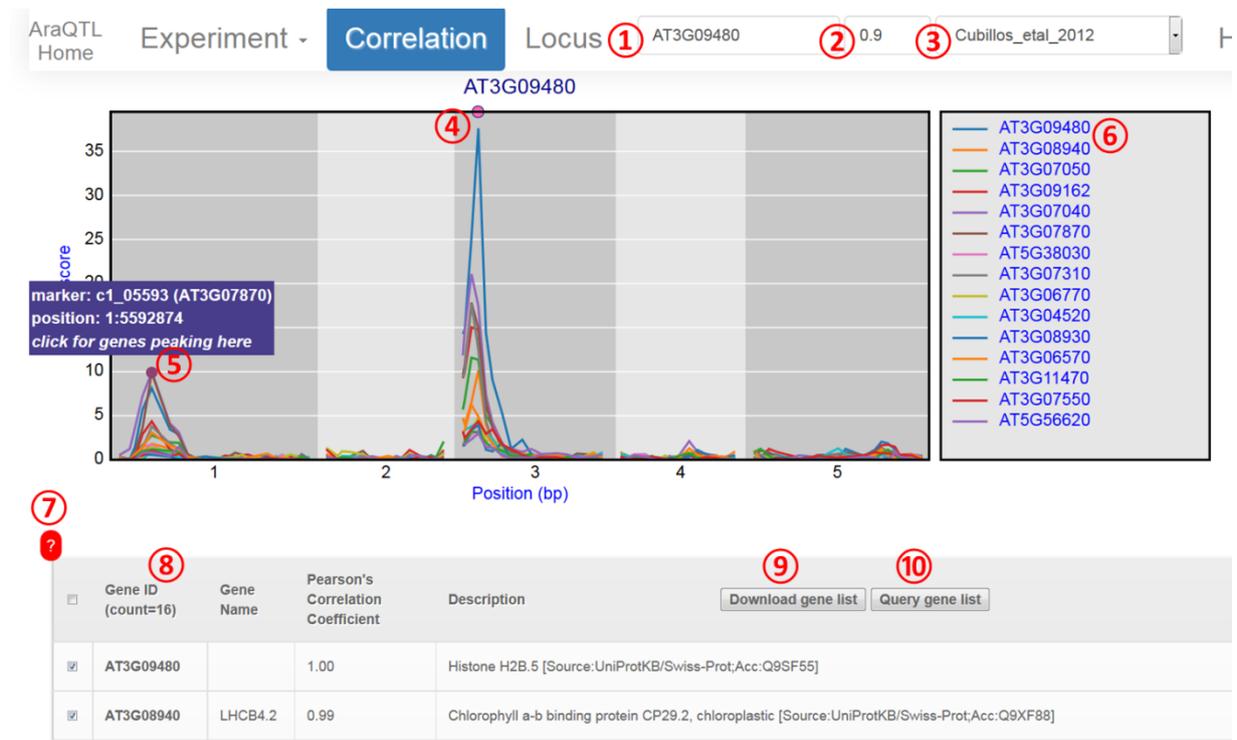


Figure 6: Correlation between eQTL profiles. The gene selection box (1), correlation threshold (2), and experiment (3) can be found in the navigation bar. The genomic position of the selected gene can be found at the upper part of the figure (4) as a pink dot. When hovered over the eQTL profile (5) extra information about the position, marker and gene is shown. The eQTL profile can be clicked to visualise all genes with an eQTL at the selected position. The gene IDs in the legend (6) can be hovered-over to highlight a specific eQTL profile and are linked to TAIR for extra information. The gene list (7) profiles selection of genes for visualization. The gene IDs in this list (8) can be clicked to go to the eQTL profiles of that gene. The gene list can be downloaded (9) for further investigation or send to the query box (10) so the eQTL profiles of the selected genes can be studied in other experiment.